Single Cell Mouse Hematopoietic Stem and Progenitor Cell Differentiation

Group: Kevin Benac, Boying Gong, George Shan, and Tae Won Chung

Diffusion Map

```r
library(destiny)
# library(plot3D)
library(dynamicTreeCut)
library(RColorBrewer)
library(slingshot)
```

```
## Loading required package: Biobase

## Loading required package: BiocGenerics

## Loading required package: methods

## Loading required package: parallel

##
## Attaching package: 'BiocGenerics'

## The following objects are masked from 'package:parallel':
##
##     clusterApply, clusterApplyLB, clusterCall, clusterEvalQ,
##     clusterExport, clusterMap, parApply, parCapply, parLapply,
##     parLapplyLB, parRapply, parSapply, parSapplyLB

## The following objects are masked from 'package:stats':
##
##     IQR, mad, sd, var, xtabs

## The following objects are masked from 'package:base':
##
##     anyDuplicated, append, as.data.frame, cbind, colMeans,
##     colnames, colSums, do.call, duplicated, eval, evalq, Filter,
##     Find, get, grep, grepl, intersect, is.unsorted, lapply,
##     lengths, Map, mapply, match, mget, order, paste, pmax,
##     pmax.int, pmin, pmin.int, Position, rank, rbind, Reduce,
##     rowMeans, rownames, rowSums, sapply, setdiff, sort, table,
##     tapply, union, unique, unsplit, which, which.max, which.min

## Welcome to Bioconductor
##
##     Vignettes contain introductory material; view with
##     'browseVignettes()'. To cite Bioconductor, see
##     'citation("Biobase")', and for packages 'citation("pkgname")'.
```

```
## Loading required package: princurve
```

```
data <- as.matrix(read.table("data/normalisedCountsVariableGenes.txt", header=TRUE, sep = ""))
dim(data) # Verify this is 4773 by 1656
```

```
## [1] 4773 1656
```

```
data_log <- log2(data+1)

# Part 1.
dist_spearman <- as.dist((1-cor(data, method="spearman"))/2)
clusters<-hclust(d=dist_spearman, method="average")
clusters_discrete<-cutreeDynamic(clusters, method="hybrid", distM=as.matrix(dist_spearman), minClusterS:
```

```
##  ..cutHeight not given, setting it to 0.3  ===>  99% of the (truncated) height range in dendro.
##  ..done.
```

```
# Part 2.
diffmap <- DiffusionMap(data=t(data_log), sigma=0.16, distance = "cosine") # This will take some time
# scatter3D(x=diffmap$DC3, y=diffmap$DC1, z=diffmap$DC2, theta = -110, phi=20, xlab="DC3", ylab="DC1", :
#           colvar=as.numeric(clusters_discrete), col=c("black","purple","magenta","cyan","yellow","blu
```

Pseudotimes

Notice that there are only two unclustered (cluster=0), so we ignore the third #lineage. Lineage 1 is the lineage for differentiation into white blood cells. Lineage 2 is the lineage for differentiation into red blood cells and Megakaryocytes. Megakaryocytes are rare, so most of lineage 2 is probably pre-erythrocyte (rbc). RBC's are weird in that they lack nuclei and are bascially carriers for hemoglobin, #which means that alot of significant gene expression had to have occured to cause this #cell fate. This is consistent with seeing how DC1 is basically defined by cluster 2 #being apart from the others
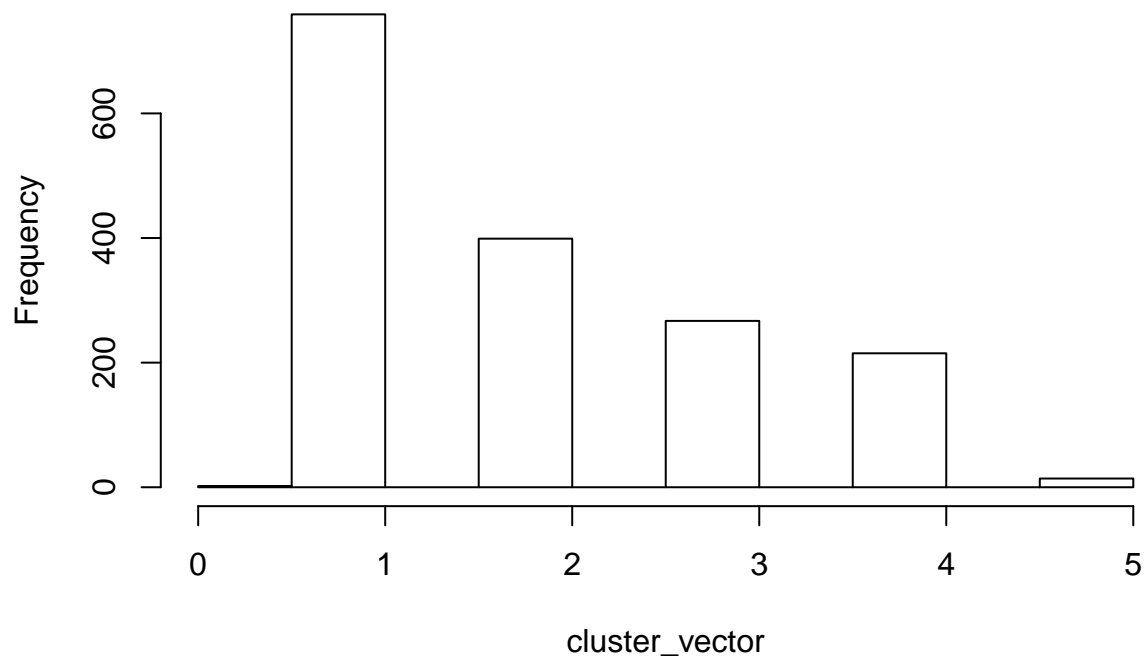
```
cluster_vector<-as.vector(clusters_discrete)
which(cluster_vector==0)
```

```
## [1] 1096 1115
```

```
hist(cluster_vector)
```

Histogram of cluster_vector
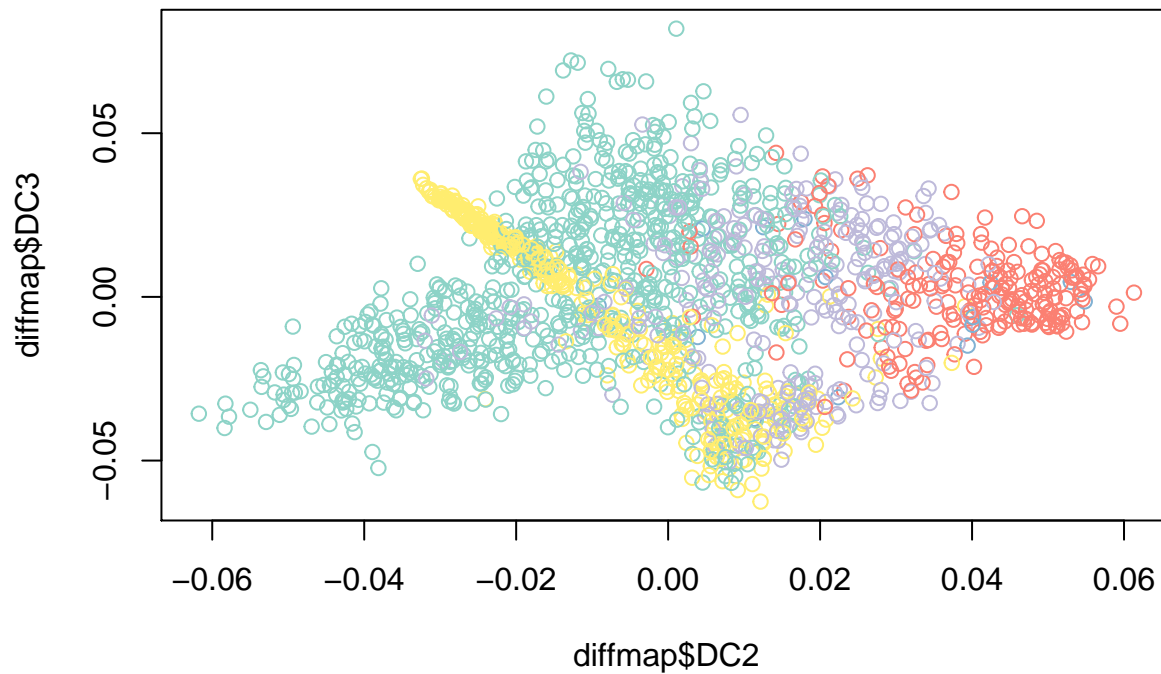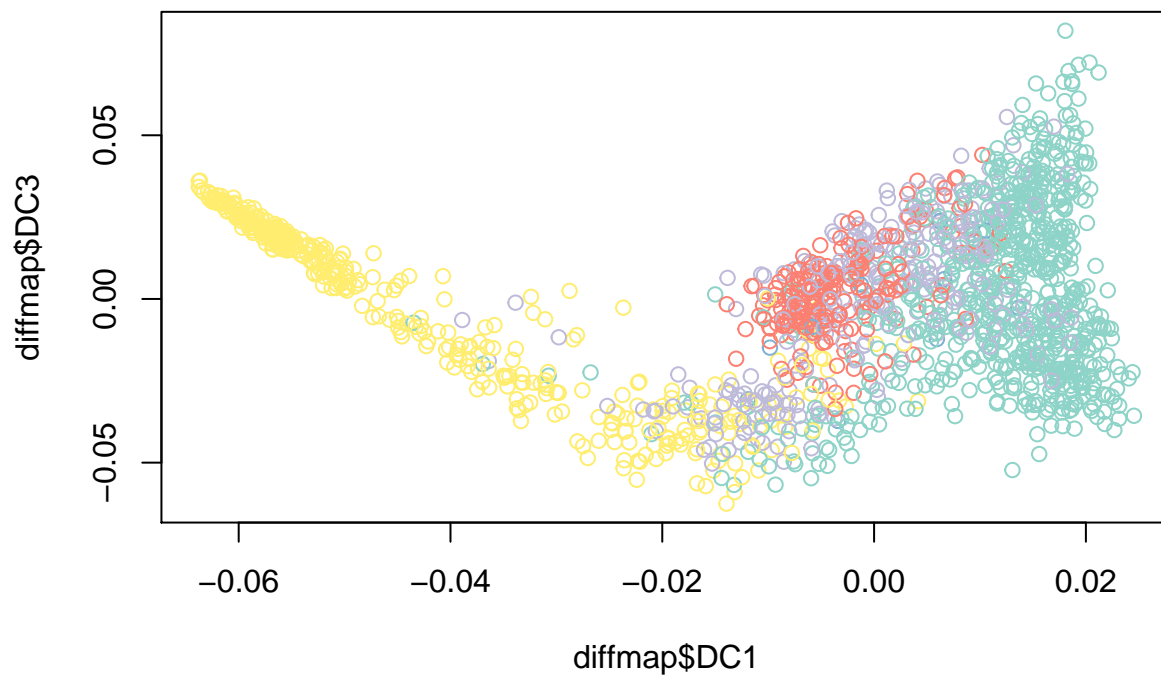


```r
rd<-cbind(diffmap$DC1, diffmap$DC2, diffmap$DC3)

plot(diffmap$DC1,diffmap$DC2, col=cluster_vector)
```
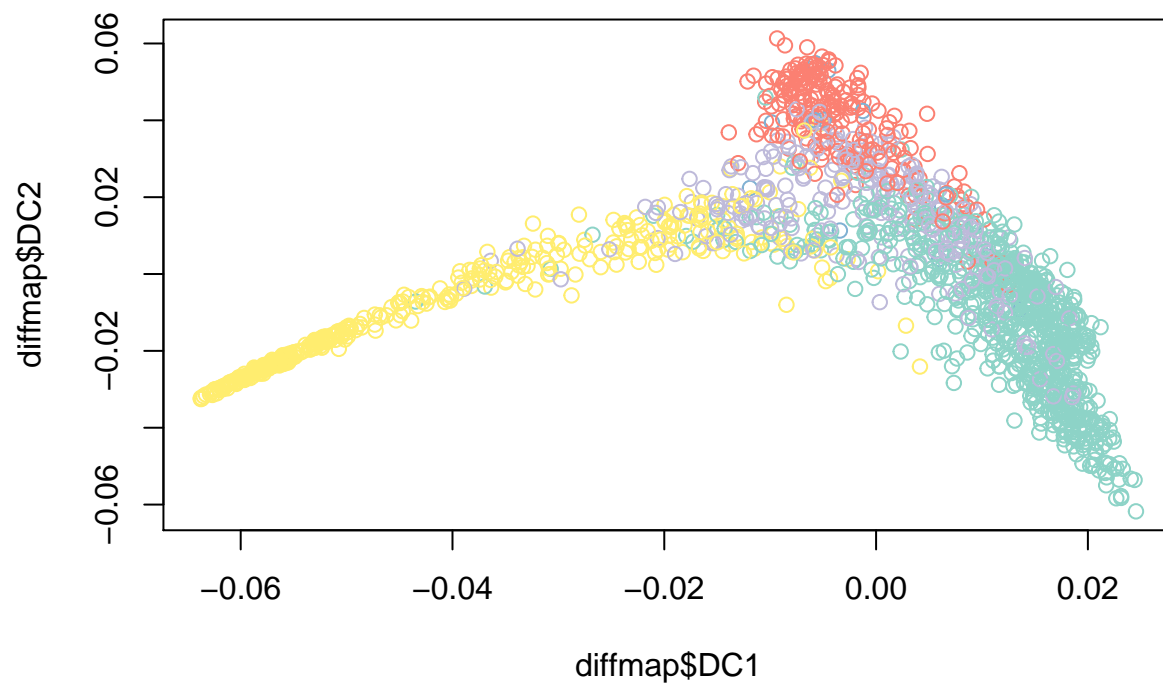


```r
plot(diffmap$DC2,diffmap$DC3, col=cluster_vector)
```

```
plot(diffmap$DC1,diffmap$DC3, col=cluster_vector)
```



```
sds<-slingshot(rd, cluster_vector, start.clus=1)
```

```
## Using diagonal covariance matrix
```

```
sds
```

```
## class: SlingshotDataSet
```

```
## 
##   Samples Dimensions
##      1656           3
## 
## lineages: 3
## Lineage1: 1  3  5  4
## Lineage2: 1  3  2
## Lineage3: 1  0
## 
## curves: 3
## Curve1: Length: 0.13876  Samples: 822.87
## Curve2: Length: 0.20971  Samples: 976.34
## Curve3: Length: 0.13603  Samples: 609.92
```

```r
psdt<-pseudotime(sds)
```

```r
head(cbind(psdt,cluster_vector), 100)
```

```
##               curve1      curve2      curve3 cluster_vector
## Cell-1    0.04765777  0.04775829          NA              3
## Cell-2    0.05351497  0.05113183  0.05211391              1
## Cell-3    0.03223960  0.03221008  0.03233601              1
## Cell-4    0.06623187  0.06150331  0.10255115              1
## Cell-5    0.06792297  0.06549150  0.09061651              1
## Cell-6    0.05887061  0.05444123  0.05662301              1
## Cell-7    0.04382467  0.04159929  0.04297008              1
## Cell-8    0.07225748  0.07982840  0.12206336              1
## Cell-9    0.03153914  0.03164424          NA              3
## Cell-10   0.00000000  0.00000000  0.00000000              1
## Cell-11   0.06200211  0.05944892  0.06159530              1
## Cell-12   0.07125159  0.07198024  0.08855529              1
## Cell-13   0.03881832  0.03966941  0.03867846              1
## Cell-14   0.04384412  0.04181709  0.04305120              1
## Cell-15   0.07649336  0.07308640  0.07105777              1
## Cell-16   0.10057401  0.05406095          NA              3
## Cell-17   0.02139675  0.02143292  0.02150561              1
## Cell-18   0.07919309  0.08258857  0.07509610              1
## Cell-19   0.11285760          NA          NA              4
## Cell-20   0.03125630  0.03129743  0.03087749              1
## Cell-21   0.08276955  0.08900068          NA              3
## Cell-22   0.03083715  0.03083437          NA              3
## Cell-23   0.11279331  0.06490648          NA              3
## Cell-24   0.02643334  0.02646441  0.02656401              1
## Cell-25   0.07085666          NA  0.10479988              1
## Cell-26   0.03887090  0.03874093  0.03881154              1
## Cell-27   0.01792333  0.01789727          NA              3
## Cell-28   0.06741237  0.06433865  0.09659694              1
## Cell-29   0.03395041  0.03390828  0.03397351              1
## Cell-30   0.06791750  0.06555054  0.08724415              1
## Cell-31   0.03868308  0.03826946  0.03866266              1
## Cell-32   0.03681710  0.03743749  0.03652017              1
## Cell-33   0.05922492  0.05864560  0.05910310              1
## Cell-34   0.05712099  0.05204938  0.05415246              1
## Cell-35   0.01087548  0.01087170  0.01086525              1
```

```
## Cell-36  0.10950309 0.06068250          NA          3
## Cell-37  0.02588606 0.02587065 0.02582458          1
## Cell-38  0.04493061 0.04368059 0.04437288          1
## Cell-39  0.03059680 0.03063040          NA          3
## Cell-40  0.13876126          NA          NA          4
## Cell-41  0.10189277          NA          NA          4
## Cell-42  0.12823001          NA          NA          4
## Cell-43          NA 0.17935945          NA          2
## Cell-44  0.08689648 0.08497288 0.07165452          1
## Cell-45          NA 0.11683090          NA          2
## Cell-46  0.13370967          NA          NA          4
## Cell-47          NA 0.20615673          NA          2
## Cell-48          NA 0.18952304          NA          2
## Cell-49          NA 0.14773131          NA          2
## Cell-50  0.09654595 0.06342065          NA          3
## Cell-51          NA 0.18741014          NA          2
## Cell-52          NA 0.16963732          NA          2
## Cell-53          NA 0.13812580          NA          2
## Cell-54          NA 0.18714437          NA          2
## Cell-55  0.11255758 0.03983579          NA          3
## Cell-56          NA 0.18845858          NA          2
## Cell-57          NA 0.09849427          NA          2
## Cell-58          NA 0.18597380          NA          2
## Cell-59          NA 0.20097275          NA          2
## Cell-60  0.07020584          NA 0.10254237          1
## Cell-61          NA 0.12892613          NA          2
## Cell-62  0.10097316 0.07981943          NA          3
## Cell-63          NA 0.18111143          NA          2
## Cell-64          NA 0.18661763          NA          2
## Cell-65          NA 0.14395494          NA          2
## Cell-66  0.05654897 0.05680090 0.05678926          1
## Cell-67  0.11413799 0.04375791          NA          3
## Cell-68          NA 0.10944691          NA          2
## Cell-69          NA 0.17679361          NA          2
## Cell-70  0.13287095          NA          NA          4
## Cell-71          NA 0.10637407          NA          2
## Cell-72          NA 0.20247441          NA          2
## Cell-73          NA 0.18875428          NA          2
## Cell-74  0.07098004          NA 0.10599399          1
## Cell-75  0.09805135 0.06450495          NA          3
## Cell-76          NA 0.19737547          NA          2
## Cell-77          NA 0.12378200          NA          2
## Cell-78          NA 0.11366224          NA          2
## Cell-79          NA 0.15403668          NA          2
## Cell-80          NA 0.14767172          NA          2
## Cell-81          NA 0.20700646          NA          2
## Cell-82  0.04218191 0.04221289 0.04211215          1
## Cell-83  0.09611015 0.06513640          NA          3
## Cell-84  0.04503997 0.04630934          NA          3
## Cell-85  0.04576554 0.04607272 0.04581396          1
## Cell-86  0.04405785 0.04411617 0.04400366          1
## Cell-87  0.09062525 0.07795360 0.06782521          1
## Cell-88  0.06858132          NA 0.10172909          1
## Cell-89  0.04673783 0.04819490 0.04732474          1
```

```
## Cell-90  0.03232529 0.03249416        NA        3
## Cell-91  0.03703299 0.03714553 0.03692169        1
## Cell-92  0.08306471 0.08641024        NA        3
## Cell-93  0.04029330 0.04088864 0.04027378        1
## Cell-94  0.07110473 0.07204589 0.09149883        1
## Cell-95  0.03931608 0.03975525 0.03922018        1
## Cell-96  0.06699586 0.06287899 0.10490671        1
## Cell-97  0.13031735        NA        NA        4
## Cell-98  0.03230470 0.03218475 0.03267832        1
## Cell-99  0.07220254 0.07807610 0.11426088        1
## Cell-100 0.13452037        NA        NA        4
```

```r
save(psdt, cluster_vector, file="output/pseudotimes.RData")
```